



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/823,685	04/14/2004	Randall J. Calistri-Yeh	55653-017	5370

7590 04/13/2006

McDermott, Will & Emery  
600 13th Street, N.W.  
Washington, DC 20005-3096

EXAMINER

BUSS, BENJAMIN J

ART UNIT PAPER NUMBER

2129

DATE MAILED: 04/13/2006

Please find below and/or attached an Office communication concerning this application or proceeding.



Art Unit: 2129

### DETAILED ACTION

Claims 53-59, 65, and 71-75 are pending in this application.

#### *Priority*

- 5 Examiner acknowledges Applicant's claim for priority based on 09/562,916 filed on 5/2/2000 and 60/177,654 filed on 1/27/2000.

#### *Information Disclosure Statement*

- 10 The information disclosure statement filed 4/14/2004 fails to comply with 37 CFR 1.98(a)(2), which requires a legible copy of each cited foreign patent document; each non-patent literature publication or that portion which caused it to be listed; and all other information or that portion which caused it to be listed. It has been placed in the application file, but the information referred to therein has not been considered.

#### *Claim Rejections - 35 USC § 112*

- 15 The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

- 20 Claims 53, 55, 57, 65, 71, 72-75 rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. It is not clear how the "trainable semantic vectors" of the instant invention are "trainable".
- 25 Neither the claims nor the specification define how to train the semantic vectors. The disclosure does not enable the person of ordinary skill in the art to make and/or use these claimed invention. Therefore, for examination of the instant application, the broadest reasonable interpretation of the phrase "trainable semantic vector" has been determined to be any vector representation which takes semantic information into consideration, since "trainable" is simply an identifying label absent a clear definition.

***Claim Rejections - 35 USC § 112***

The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claim 57 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 57 recites the limitation "the trainable semantic vectors of each of the data points in a semantic lexicon" in lines 3-4 of the claim. There is insufficient antecedent basis for this limitation in the claim.

***Claim Rejections - 35 USC § 101***

35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 53-59, 65, and 71-75 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. The computer system must set forth a practical application of a §101 judicial exception to produce a real-world result. Benson, 409 U.S. at 71-72, 175 USPQ at 676-77. The invention is ineligible because it has not been limited to a substantial practical application. An invention directed to selecting abstract datasets or data points is useless in a real world situation. Further adding of selected abstract data points to an abstract dataset is also useless in a real world situation.

In determining whether the claim is for a "practical application," the focus is not on whether the steps taken to achieve a particular result are useful, tangible, and concrete. If the claim is directed to a practical application of the §101 judicial exceptions producing a result tied to the physical world that does not preempt the judicial exception, then the claim meets the statutory requirement of 35 U.S.C. §101.

The phrases 'constructing a trainable semantic vector', 'receiving a query containing information', 'comparing the trainable semantic vector for the query to the trainable semantic vector of each dataset', and 'selecting datasets whose trainable semantic vectors are closest to the trainable semantic vector for the query' are a manipulation of abstract concepts and are not clear in purpose or scope. Other variations on these phrases in the claims, as well as further 'adding said selected data points to said dataset', do not provide a clear purpose or scope for the claimed invention.

Art Unit: 2129

The invention must be for a practical application and either:

- 1) specify transforming (physical thing – article) or
- 2) have the Final Result (not the steps) achieve or produce a  
useful (specific, substantial, AND credible),  
concrete (substantially repeatable/non-unpredictable), AND  
tangible (real world/non-abstract) result  
(tangibility is the opposite of abstractness).

A claim that is so broad that it reads on both statutory and non-statutory subject matter must be amended, and if the specification discloses a practical application but the claim is broader than the disclosure such that it does not require the practical application, then the claim must be amended.

Claims that construct data representations, receive queries, construct further data representations, compare data representations, and select data based on similarities between representations are not statutory.

Claims further combining data are still not statutory.

Appropriate corrections are required.

### ***Claim Rejections - 35 USC § 102***

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

Claims 53-59, 65, and 71 are rejected under 35 U.S.C. 102(e) as being anticipated by Liddy (USPN 5,963,940).

#### **Claim 53:**

Liddy anticipates:

- Constructing a trainable semantic vector for each dataset (C2:37 especially "documents in the database have preferably been processed to provide corresponding alternative representations" C2:60-65 and

Art Unit: 2129

- 95 "semantic" C6:60-C7:7 and "Natural Language Processing" and "subject-based vector representation of the document's contents" C9:30-60; Also see the Abstract);
- Receiving a query containing information indicative of desired datasets (C2-37 especially "query" C2:45-C3:20 and C12:50-60; Also see the Abstract);
  - Constructing a trainable semantic vector for the query (C2-37 especially "processes the query to generate an alternative representation ... is a subject field code vector" C2:45-67 and "texts (document and queries) are processed to determine discourse aspects of the text" C3:20-35 and "semantic" C6:60-C7:7 and "Natural Language Processing" and "subject-based vector representation of the document's contents" C9:30-60 and "Query Processing" C16:48-C17:10; Also see the Abstract);
  - Comparing the trainable semantic vector for the query to the trainable semantic vector of each dataset (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 and "query representation is matched to the relevant document database" C2:55-65); and
  - Selecting datasets whose trainable semantic vectors are closest to the trainable semantic vector for the query (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 and "displays documents judged relevant to the content of the query" C3:40-60).

**Claim 54:**

Liddy anticipates:

- 115 - Wherein the datasets correspond to documents and the query is a natural language query (C2-37 especially "natural language query" C2:48-C3:15 and C5:1-7).

**Claim 55:**

Liddy anticipates:

- 120 - Performing a second search for datasets within the collection of datasets, wherein the second search using a method other than trainable semantic vectors (C2-37 especially "proper noun matching" C24:64-:25:15 and "Complex Nominals ... Mandatory requirements" C21:25-55);

Art Unit: 2129

- Combining the two search results to obtain a combined weighted score for each dataset in either of the two search results (C2-37 especially "Weighted Boolean Processor" C20:54-C21:20 and "Scoring" C22:1-8);
- 125 - Selecting datasets whose combined weighted score is largest (C2-37 especially "highest weighted score" C21:1-13).

**Claim 56:**

Liddy anticipates:

- 130 - Further comprising a step of clustering the selected datasets in real time (C2-37 especially "real time" C7:20-35).

**Claim 57:**

Liddy anticipates:

- 135 - Constructing a trainable semantic vector for the dataset (C2-37 especially "documents in the database have preferably been processed to provide corresponding alternative representations" C2:60-65 and "semantic" C6:60-C7:7 and "Natural Language Processing" and "subject-based vector representation of the document's contents" C9:30-60; Also see the Abstract);
- Comparing the trainable semantic vector for the dataset to the trainable semantic vectors of each of the  
140 data points in a semantic lexicon (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 and "query representation is matched to the relevant document database" C2:55-65 and "lexicon" C9:50-60 and C17:15-35);
- Selecting data points whose trainable semantic vectors are closest to the trainable semantic vector for the  
145 dataset (C2-37 especially "The matching score between the query and document is determined ... computes the distance between these two data points" C23:20-35 and "displays documents judged relevant to the content of the query" C3:40-60);
- Adding said selected data points to said dataset (C2-37 especially "The vector for the original query and ...  
150 documents are weighted and combined to form a new, single vector for re-ranking and re-Clustering" C26:10-17).

Art Unit: 2129

**Claim 58:**

**Liddy** anticipates:

- Wherein the dataset is a document and the data points are words (C2-37 especially "all words in the document" C9:50-60 and "documents ... words" C15:24-32 and "word" and "document" C16:1-25).

**Claim 59:**

**Liddy** anticipates:

- Wherein the dataset is a natural language query in a search system and the data points are words (C2-37 especially "natural language ... words" C15:24-32).

**Claims 65 and 71:**

Claims 65 and 71 are rejected on the same grounds as claim 53 above since the recited limitations are substantially the same.

***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this

Office action:

- (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

- Claims 72-75 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Liddy** (USPN 5,963,940) in view of **Hazlehurst** (USPN 5,974,412).

**Claim 72:**

**Liddy** teaches:

- The query or each of the datasets includes at least one data point (C2-37 especially "words" C9:50-60 and C15:24-32 and C16:1-25 and "data points" C23:20-35); and



Art Unit: 2129

- The trainable semantic vector for the query or each of the datasets is constructed by the steps of:
  - o Determining the significance of each data point with respect to the predetermined categories (C2-37 especially "Each information bearing word in a text is looked up in the online, lexical resource ... representation of the document's contents" C9:50-60);
  - o Constructing a trainable semantic vector for each data point, wherein each trainable semantic vector has dimensions equal to the number of predetermined categories and represents the relative strength of its corresponding data point with respect to each of the predetermined categories (C2-37 especially "vector representation" C9:30-60); and
  - o Combining the trainable semantic vector for each of the at least one data point to form the semantic vector of the query or each of the datasets (C2-37 especially "The vector for the original query and ... documents are weighted and combined to form a new, single vector for re-ranking and re-Clustering" C26:10-17 and "vector representation" C9:30-60 and "" C).

**Liddy** fails to teach:

- The step in constructing the trainable semantic vector for the query or each of the datasets of:
  - o For each data point, constructing a table for storing information indicative of a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space.

**Hazlehurst** teaches:

- The step in constructing the trainable semantic vector for the query or each of the datasets of:
  - o For each data point, constructing a table for storing information indicative of a relationship between each data point and predetermined categories corresponding to dimensions in the semantic space (C2-28 especially "vector spaces utilized for the categorization and retrieval of documents" C6:10-25 and "table listing features ... significant markers of semantic content" C9:20-41 and "table ... semantically related documents ... centroid table ... for each centroid vector output .... vector space ... document table ... document vectors ... clustering ... semantic similarity ... document vector and each centroid vector" C11:65-C12:50).

Motivation:

**Liddy** and **Hazlehurst** are from the same field of endeavor, information retrieval. It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the teachings of **Liddy** by

Art Unit: 2129

210 constructing a table for each data point that stores information indicative of a relationship between each data point and categories corresponding to the semantic vector space as taught by **Hazlehurst** for the benefit of improved efficiency and performance at categorizing and retrieving documents (**Hazlehurst** C6:65-C7:8), to facilitate user queries (**Hazlehurst** C4:33-38), and to relate centroid vectors to closest document vectors and document vectors to closest centroid vectors (**Hazlehurst** C12:20-40)

215

**Claims 73, 74, and 75:**

Claims 73-75 are rejected on the same grounds as claim 72 above since the recited limitations are substantially the same.

220

**Conclusion**

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

225

- Kuromusha (USPN 6,175,828)
- Gillis (USPN 6,523,026)
- Noyes (USPN 5,379,366)
- Fujisawa (USPN 5,404,506)
- Schaefer (USPN 5,826,268)
- Punch (USPN 5,924,105)
- Kitagawa (USPN 6,138,116)
- Bellegarda (USPN 6,208,971)
- 230 - Carter (USPN 6,405,199)
- Bartell ("Representing Documents Using an Explicit Model of Their Similarities")
- Bullinaria ("Modelling Lexical decision Using Corpus Derived Semantic Representations in a Connectionist Network")
- Lewis ("Text Representation for Intelligent Text Retrieval: A Classification-Oriented View")
- 235 - Liu ("Semantic Vector Space Model: Implementation and Evaluation")
- Liu ("The Semantic Vector Space Model (SVSM): A Text Representation and Searching Technique")
- Patel ("Extracting Semantic Representations from Large Text Corpora")

Art Unit: 2129

Claims 53-59, 65, and 71-75 are rejected.

240

***Correspondence Information***

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Benjamin J. Buss whose telephone number is 571-272-5831. The examiner can normally be reached on M-F 9AM-5PM.

245

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David Vincent can be reached on 571-272-3080. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

250

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

255 BJB

Benjamin J Buss  
Examiner  
Art Unit 2129

A handwritten signature in black ink, appearing to read "Benjamin J. Buss" with a stylized flourish at the end.